10. The SKY/CGH Database for Spectral Karyotyping and Comparative Genomic Hybridization Data

by Turid Knutsen, Vasuki Gobu, Rodger Knaus, Thomas Ried, and Karl Sirotkin

National Cancer Institute (NCI)
National Center for Biotechnology Information (NCBI)

Summary

Spectral Karyotyping (SKY) (1–7) and Comparative Genomic Hybidization (CGH) (8–11) are complementary fluorescent molecular cytogenetic techniques that have revolutionized the detection of chromosomal abnormalities. SKY permits the simultaneous visualization of all human or mouse chromosomes in a different color, facilitating the detection of chromosomal translocations and rearrangements (Figure 1). CGH uses the hybridization of differentially labeled tumor and reference DNA to generate a map of DNA copy number changes in tumor genomes.

Figure 1: A metaphase spread after SKY hybridization.
The RGB image demonstrates cytogenetic abnormalities in a cell from a secondary leukemia cell line. *Arrows* indicate some of the many chromosomal translocations in this cell line.

The goal of the SKY/CGH database is to allow investigators to submit and analyze both clinical and research (e.g., cell lines) SKY and CGH data. The database is growing and currently has a total of about 700 datasets, some of which are being held private until published. Several hundred labs around the world use this technique, with many more looking at the data they generate. Submitters can enter data from their own cases in either of two formats, public or private; the public data is generally that which has already been published, whereas the private data can be viewed only by the submitters, who can transfer it to the public format at their discretion. The results are stored under the name of the submitter and are listed according to case number. The homepage includes a basic description of SKY and CGH techniques and provides links to a more detailed explanation and relevant literature.

Database Content

Detailed information on how to submit data either to the SKY or CGH sectors of the database can be found through links on the homepage. What follows is a brief outline.

Spectral Karyotyping

The submitter enters the written karyotype, the number of normal and abnormal copies for each chromosome, and the number of cells for each clone. Each abnormal chromosome segment is then described by typing in the beginning and ending bands, starting from the top of the chromosome (Figure 2); the computer then builds a colored ideogram of this chromosome and eventually a full karyotype (SKYGRAM) with each normal and abnormal chromosome displayed in its unique SKY classification color, with band overlay (Figure 3). Each breakpoint submitted is automatically linked by a button marked FISH to the human Map Viewer (Figure 4; Chapter 19), which provides a list of genes at that site and available FISH clones for that breakpoint.

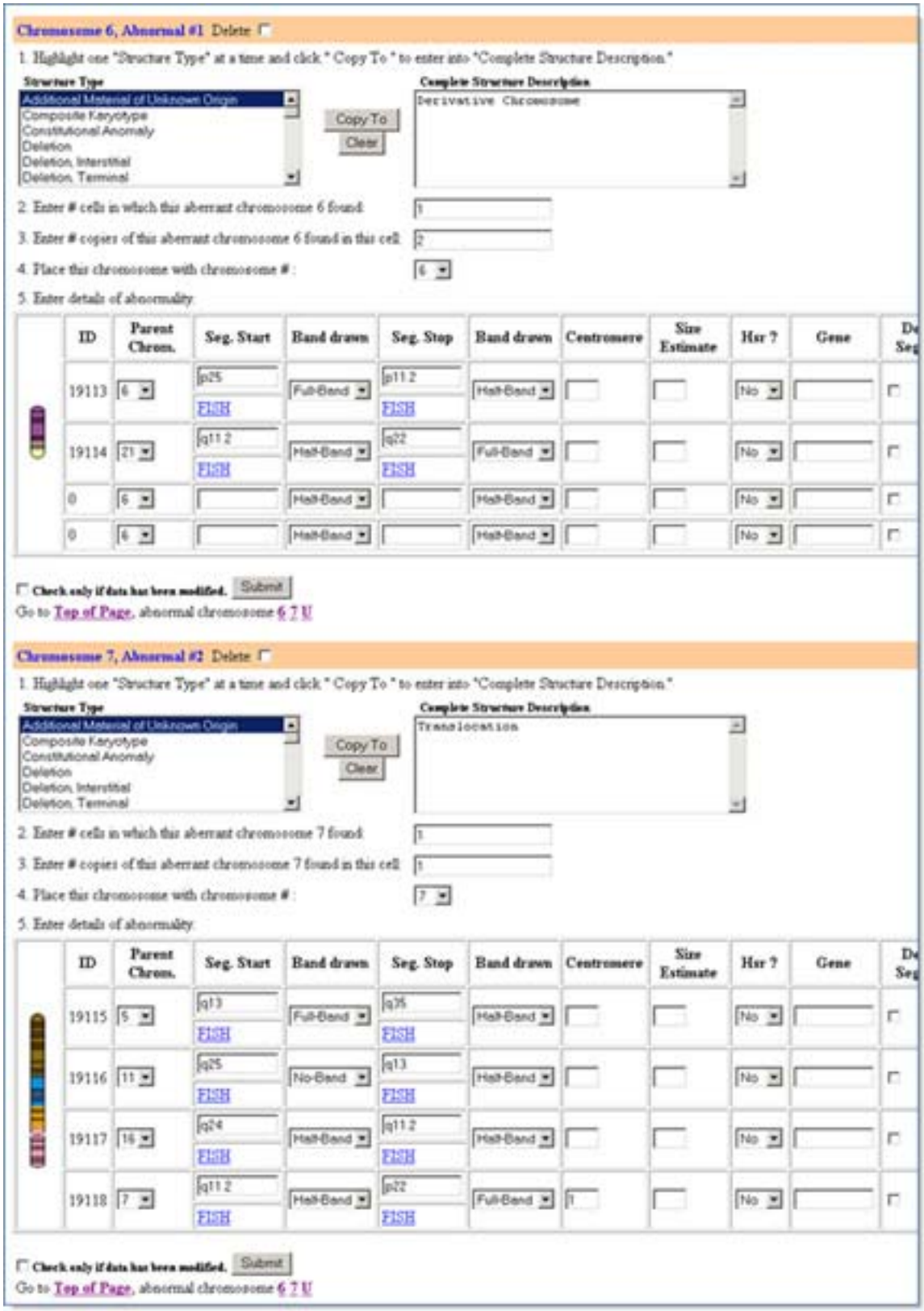

Figure 2: SKY data entry form for two different abnormal chromosomes, built segment by segment, for the SKYGRAM image.

Chromosome images on the *left* are the result of entering the start (*top*) and stop (*bottom*) band for each segment.

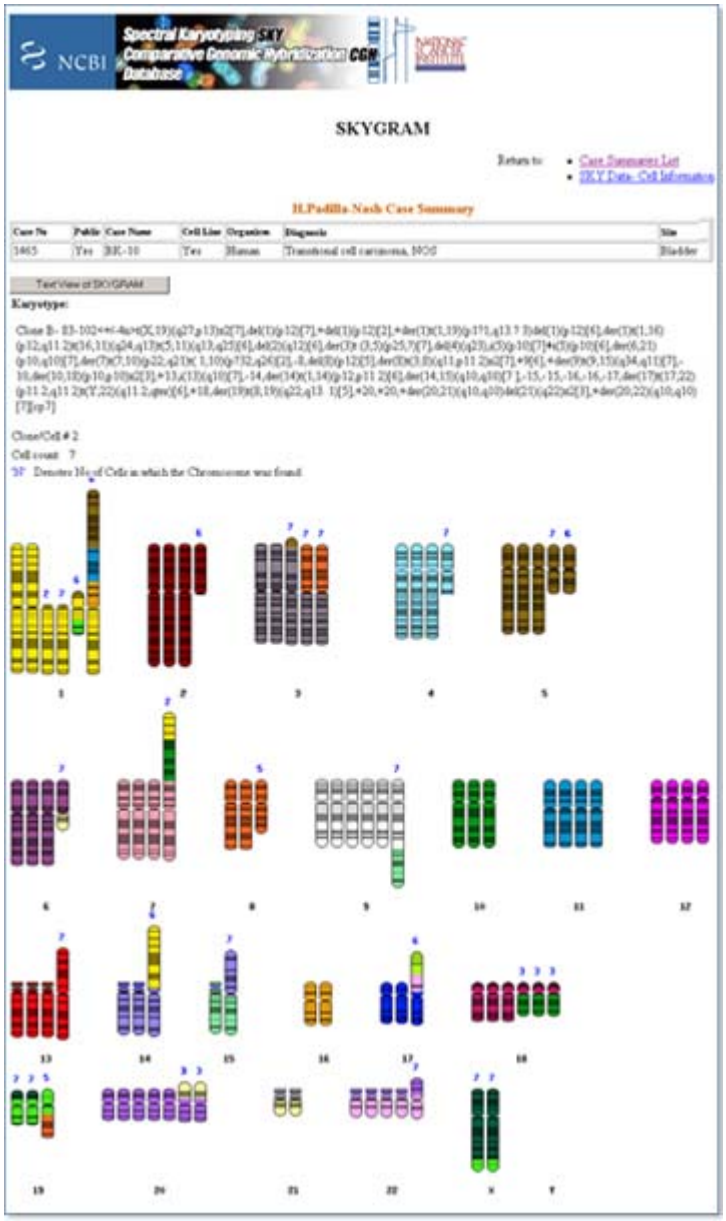

Figure 3: A SKYGRAM from the SKY/CGH database demonstrating cytogenetic abnormalities in a transitional cell carcinoma of the bladder cell line.
The written karyotype and a one-line case summary are also provided.

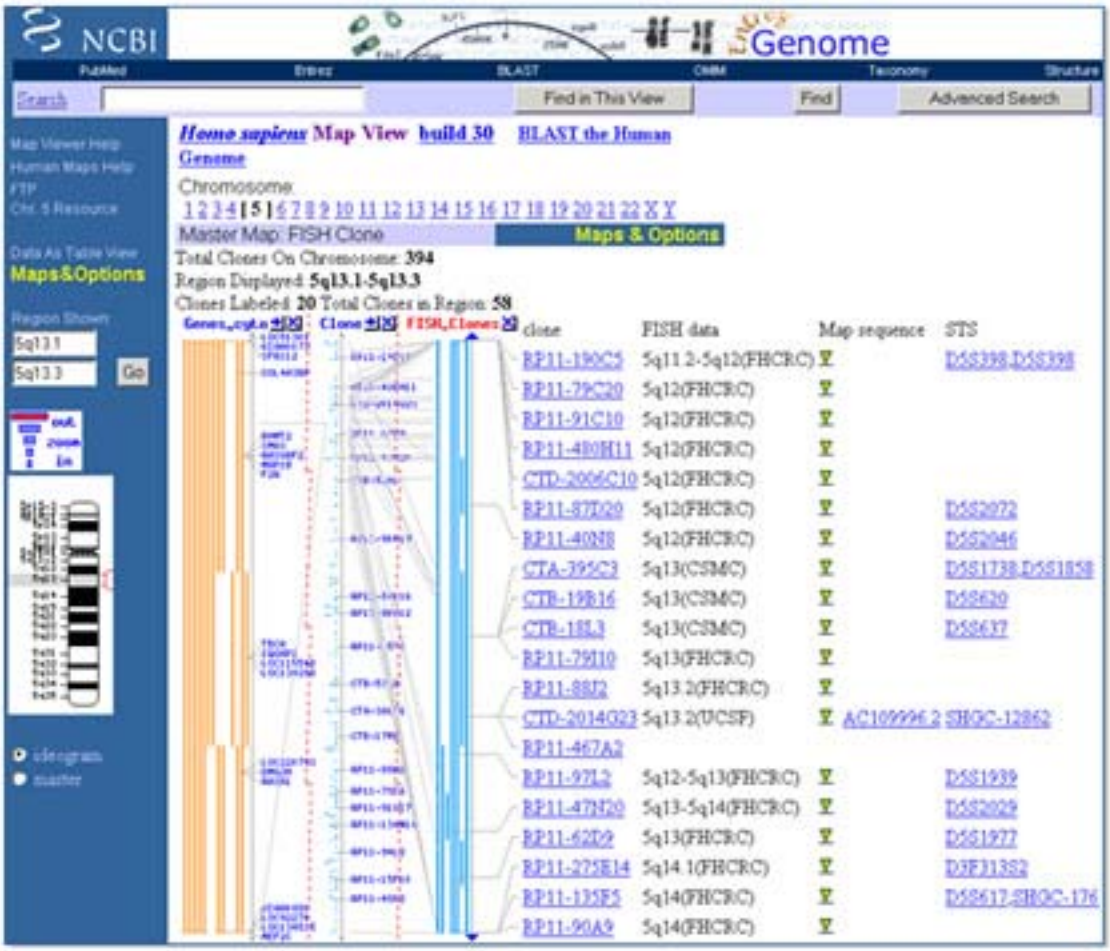

Figure 4: Map Viewer image depicting the information on genes, clones, FISH clones, map sequences, and STSs for a specific chromosomal breakpoint (5q13) identified in a SKYGRAM image.

Comparative Genomic Hybridization

The CGH database displays gains, losses, and amplification of chromosomes and chromosome segments. The data are entered either by hand or automatically from various CGH software programs. In the manual format, the submitter enters the band information for each affected chromosome, describing the start band and stop band for each gain or loss, and the computer program displays the final karyotype with vertical bars to the left or right of each chromosome, indicating loss or gain, respectively (Figure 5).

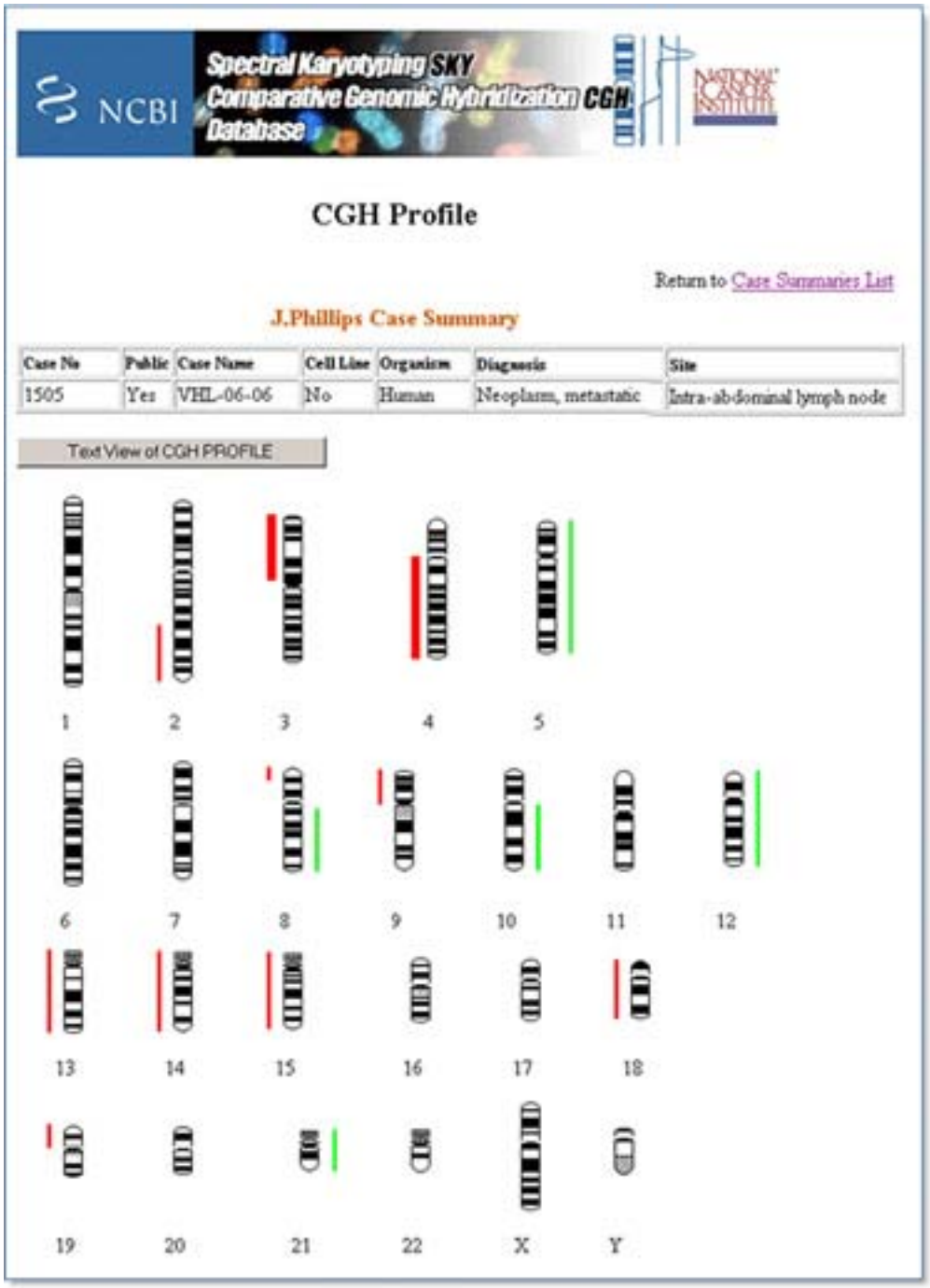

Figure 5: A CGH profile from the SKY/CGH database demonstrating copy number gains and losses in a tumor cell line established from a metastatic lymph node in von Hippel-Lindau renal cell carcinoma.

Case Information

Clinical data submitted include case identification, World Health Organization (WHO) disease classification code, diagnosis, organ, tumor type, and disease stage. To obtain the correct classification code, a link is provided to the NCI's Metathesaurus^{TM}site, which includes, among its many systems, the codes developed by the WHO and NCI, and published as the International Classification of Diseases, 3rd edition (ICD-O-3).

Reference Information

The references for the published cases are entered into the Case Information page and are linked to their abstracts in PubMed.

Tools for Data Entry

SKYIN

A colored karyotype with band overlay is presented to the submitter, who then builds each aberrant chromosome by cutting and pasting (by clicking with the mouse at appropriate breakpoints) (Figure 6). Each aberrant chromosome is then inserted into the full karyotype, which is displayed as a SKYGRAM.

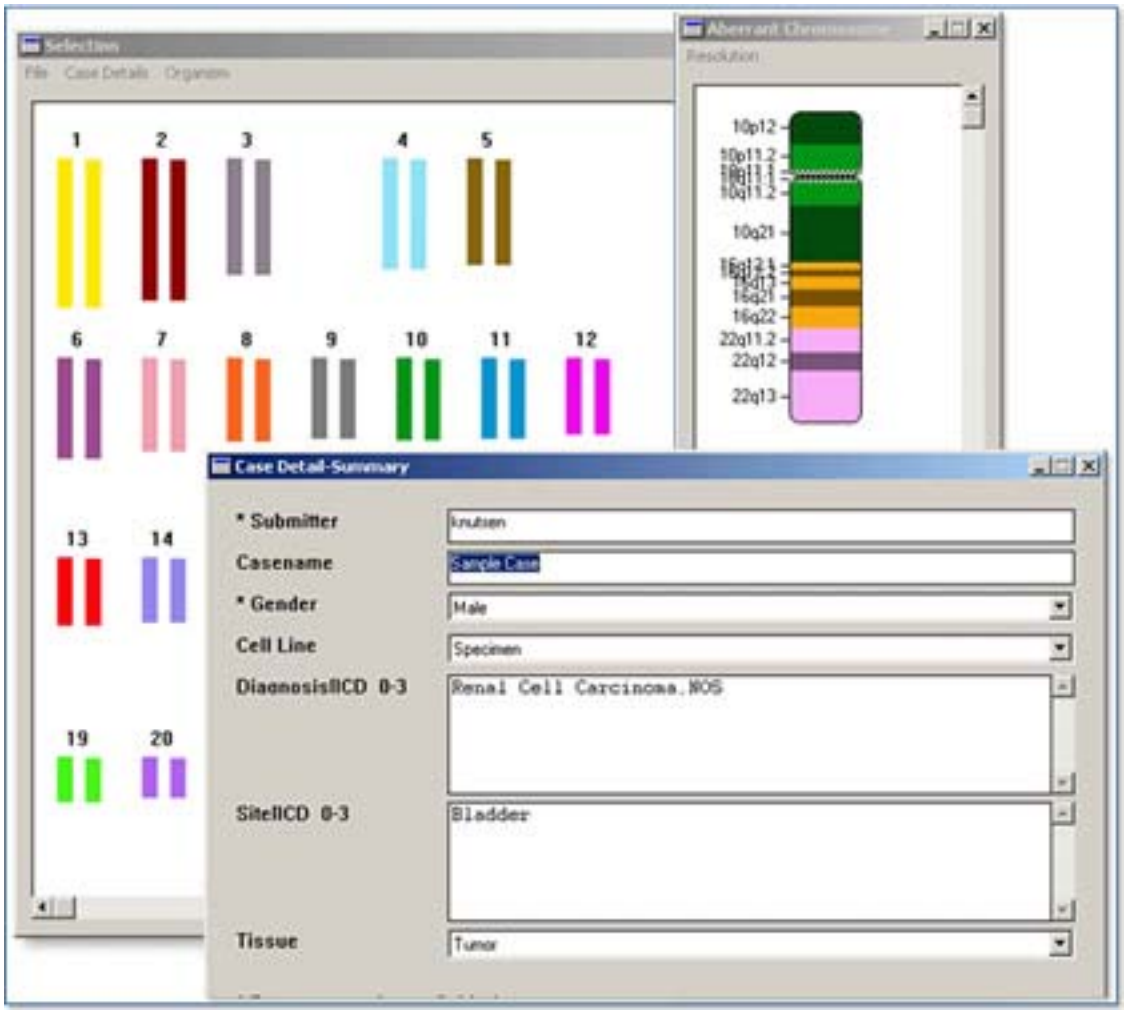

Figure 6: SKYIN format.

Clicking on a chromosome brings up that chromosome with band overlay. Using the cursor, the operator cuts and pastes together each abnormal chromosome. The abnormal chromosome shown is a combination of chromosomes 15, 16, and 4. *Inset,* an example of the clinical information entered for a case.

Karyotype Parser

To speed up the entry of cytogenetic data into the database, NCBI has built a computer program to automatically read short-form karyotypes, extract the information therein, and insert it into the SKY database (Figure 7). Karyotypes are written according to specific rules described in *An International System for Human Cytogenetic Nomenclature (1995)* (12). Using these rules, the parser (1) breaks the karyotype into small syntactic components, (2) assembles information from these components into an information structure in computer memory, (3) transforms this information into the formats required for an application, and (4) uses the information in the application, i.e., inserts it into the database. To accomplish this, the syntactic parser first extracts the information out of each piece of the input; the pieces are then put directly into a tree structure that represents karyotype semantics. For insertion into the SKY database, the karyotype information is transformed into ASN.1 structures that reflect the design of the database.

Figure 7: Karyotype Parser.
The short-form written karyotype, entered in the karyotype field in (**a**), has been converted into a modified long-form karyotype (**b**), which describes each abnormal chromosome from *top* to *bottom*. Both short and long terms use standardized symbols and abbreviations specified by ISCN.

NCI Metathesaurus

Data submitters must use the same terminology for diagnosis (morphology) and organ site (topography) to permit comparison or combination of the data in the SKY/CGH database. From the many different disease classification systems, the *International Classification of Diseases for Oncology*, 3rd edition (ICD-O-3)(13) was selected as the database's standard. It contains a morphology tree and a topography tree. In most cases, the submitter must select one term from each tree to fully classify a case. To find and select the correct ICD-O-3 morphology and topography terms, the user is referred to NCI

Metathesaurus, a comprehensive biomedical terminology database, produced by the NCI Center for Bioinformatics Enterprise Vocabulary Service. This tool facilitates mapping concepts from one vocabulary to other standard vocabularies.

Data Analysis: Query Tools

Quick Search Format

Quick Search can be found at the top of the SKY/CGH homepage and can be used for several types of information in the database; these are defined in Searchable Topics in the Help section. Topics include cytogenetic information (whole chromosome, chromosome arm, or chromosome breakpoint), submitter name, case name, cell line by name, diagnosis, site of disease, treatment, hereditary disorders, mouse strain, and genotype. One or more terms can be entered, and there are options to search SKY alone, CGH alone, SKY AND CGH, or the default, SKY OR CGH.

The query results page displays information on all relevant cases, clones, and cells, along with details of SKY and/or CGH studies and clinical information for each case.

Advanced Search Format

All of the public clinical and cytogenetic information can be searched. This format is currently under development.

The CGH Case Comparison Tool

This tool compares and summarizes the CGH profiles from multiple cases on one ideogram. There are numerous criteria that can be used for comparison, such as diagnosis, tumor site, mouse strain, and gain or loss of specific chromosomes, chromosome arms, or chromosome bands.

Data Integration

Integration with the NCBI Map Viewer

All chromosomal bands, including breakpoints involved in chromosomal abnormalities, are linked to the Map Viewer database (Figure 4 ; see also Chapter 19) The Map Viewer provides graphical displays of features on NCBI's assembly of human genomic sequence data as well as cytogenetic, genetic, physical, and radiation hybrid maps. Map features that can be seen along the sequence include NCBI contigs, the BAC tiling path, and the location of genes, STSs, FISH mapped clones, ESTs, GenomeScan models, and variation (SNPs; see Chapter 5).

SKY/CGH Database Links

Links are provided to related websites including: chromosome databases (e.g., the Mitelman database); other NCI (e.g., CGAP and CCAP) and NCBI [e.g., the Map Viewer (Chapter 19), LocusLink (Chapter 18) resources; and PubMed (Chapter 2)] sites; The Jackson Laboratory; and several other CGH sites.

Contributors

NCBI: Karl Sirotkin, Vasuki Gobu, Rodger Knaus, Joel Plotkin, Carolyn Shenmen, and Jim Ostell

NCI: Turid Knutsen, Hesed Padilla-Nash, Meena Augustus, Evelin Schröck, Ilan R. Kirsch, Susan Greenhut, James Kriebel, and Thomas Ried

References

1. Schröck E, du Manoir S, Veldman T, Schoell B, Wienberg J, Ferguson-Smith MA, Ning Y, Ledbetter DH, Bar-Am I, Soenksen D, et al. Multicolor spectral karyotyping of human chromosomes. Science 273:494–497; 1996.

2. Liyanage M, Coleman A, du Manoir S, Veldman T, McCormack S, Dickson RB, Barlow C, Wynshaw-Boris A, Janz S, Wienberg J, et al. Multicolour spectral karyotyping of mouse chromosomes. Nat Genet 14:312–315; 1996.

3. Ried T, Liyanage M, du Manoir S, Heselmeyer K, Auer G, Macville M, Schröck E. Tumor cytogenetics revisited: comparative genomic hybridization and spectral karyotyping. J Mol Med 75:801–814; 1997.

4. Weaver ZA, McCormack SJ, Liyanage M, du Manoir S, Coleman A, Schröck E, Dickson RB, Ried T. A recurring pattern of chromosomal aberrations in mammary gland tumors of MMTV–c-myc transgenic mice. Genes Chromosomes Cancer 25:251–260; 1999.

5. Knutsen T, Ried T. SKY: a comprehensive diagnostic and research tool. A review of the first 300 published cases. J Assoc Genet Technol 26:3–15; 2000.

6. Padilla-Nash HM, Heselmeyer-Haddad K, Wangsa D, Zhang H, Ghadimi BM, Macville M, Augustus M, Schröck E, Hilgenfeld E, Ried T. Jumping translocations are common in solid tumor cell lines and result in recurrent fusions of whole chromosome arms. Genes Chromosomes Cancer 30:349–363; 2001.

7. Phillips JL, Ghadimi BM, Wangsa D, Padilla-Nash H, Worrell R, Hewitt S, Walther M, Linehan WM, Klausner RD, Ried T. Molecular cytogenetic characterization of early and late renal cell carcinomas in Von Hippel-Lindau (VHL) disease. Genes Chromosomes Cancer 31:1–9; 2001.

8. Kallioniemi A, Kallioniemi OP, Sudar D, Rutovitz D, Gray JW, Waldman F, Pinkel D. Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. Science 258:818–821; 1992.

9. Heselmeyer K, Schröck E, du Manoir S, Blegen H, Shah K, Steinbeck R, Auer G, Ried T. Gain of chromosome 3q defines the transition from severe dysplasia to invasive carcinoma of the uterine cervix. Proc Natl Acad Sci U S A 93:479–484; 1996.

10. Ried T, Liyanage M, du Manoir S, Heselmeyer K, Auer G, Macville M, Schröck E. Tumor cytogenetics revisited: comparative genomic hybridization and spectral karyotyping. J Mol Med 75:801–814; 1997.

11. Forozan F, Karhu R, Kononen J, Kallioniemi A, Kallioniemi OP. Genome screening by comparative genomic hybridization. Trends Genet 13:405–409; 1997.

12. ISCN: An International System for Human Cytogenetic Nomenclature. In: Mitelman F, editor. Basel: S. Karger; 1995.

13. Fritz A, Percy C, Jack Andrew, Shanmugaratnam K, Sobin L, Parkin DM, Whelan S, editors. International Classification of Diseases for Oncology, 3rd ed. Geneva: World Health Organization; 2000.